

Caterpillar Data Analysis

Eduardo Lopez-Quezada

2024-03-08

Introduction

- I would like to explore data from an experiment that was investigating bird predation on caterpillars. The experiment placed bags over branches of trees where a caterpillar was living to prevent predation from birds. They collected various data on the trees, branches, and caterpillars of the experimental bagged group and an unbagged control group. This dataset has a total of 2948 observations and 12 variables. I would like to analyze and compare the total amount of leaf area on a branch (`totalvarea_m2`) and whether or not the branch was bagged (`birdtreatment`). I want to know if bagged branches allowed caterpillars to consume more of the total leaf area than on the unbagged branches, due to the protection provided by the bags.

Univariate Exploration

Total Leaf Area (m^2)

Statistics

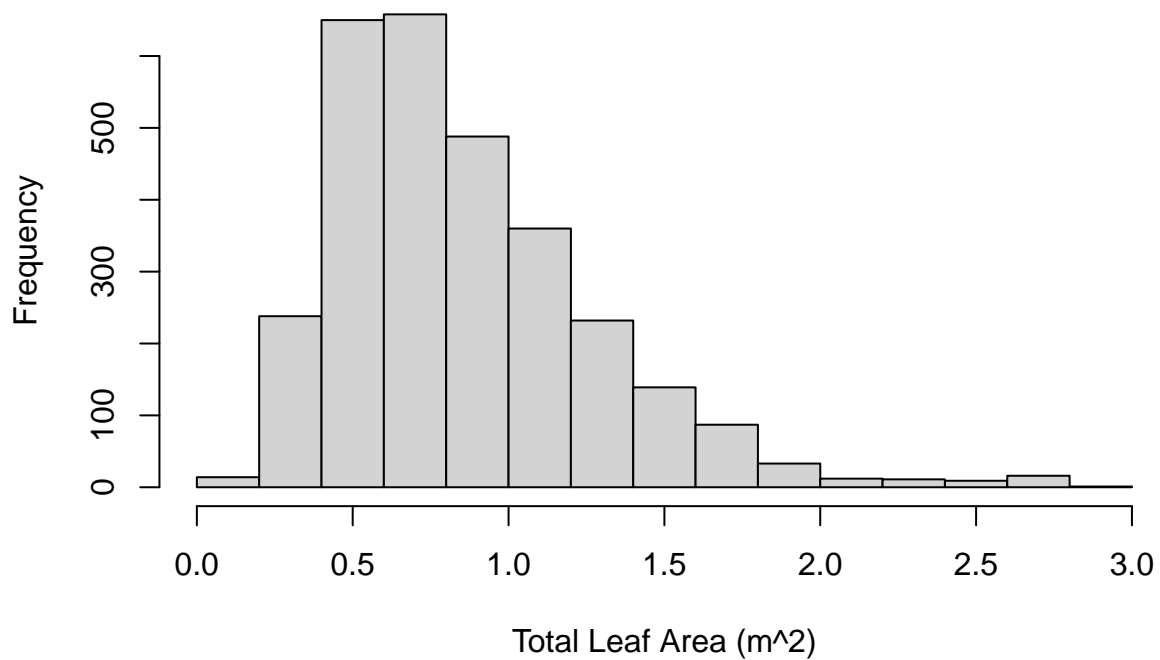
```
summary(Copy_of_Caterpillar$totallvarea_m2)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.08161 0.55725 0.76933 0.86028 1.08473 2.98759
```

Plot

```
hist(Copy_of_Caterpillar$totallvarea_m2, main="Histogram of Total Leaf Area", xlab="Total Leaf Area ( $m^2$ )")
```

Histogram of Total Leaf Area



- As shown in the histogram, the frequency of the total leaf area is skewed to the right with a center around 0.7 m². This means a majority of branches had relatively low total leaf area, but some had areas as large as 3.0 m².

Bag Status (Bagged/Unbagged)

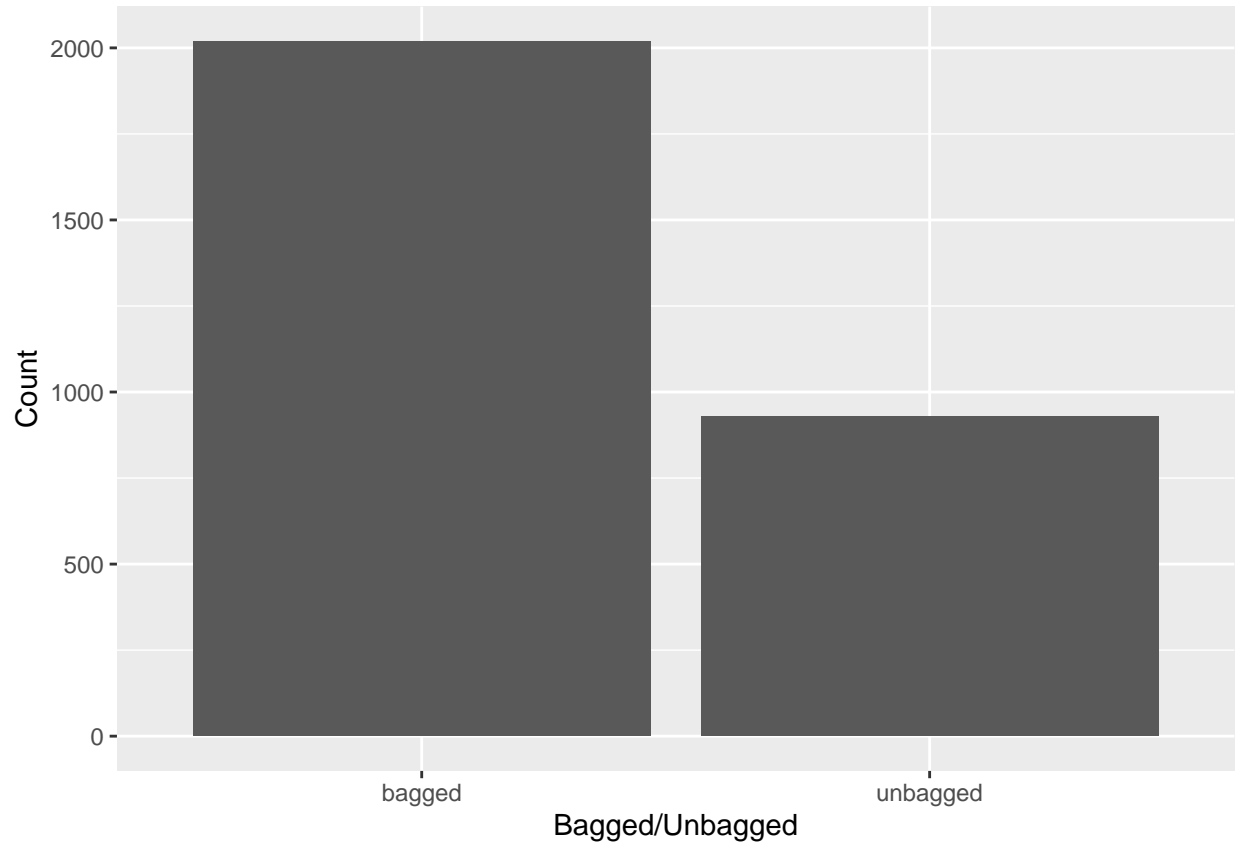
Statistics

```
table(Copy_of_Caterpillar$birdtreatment)
```

```
##  
##   bagged unbagged  
##   2019     929
```

Plot

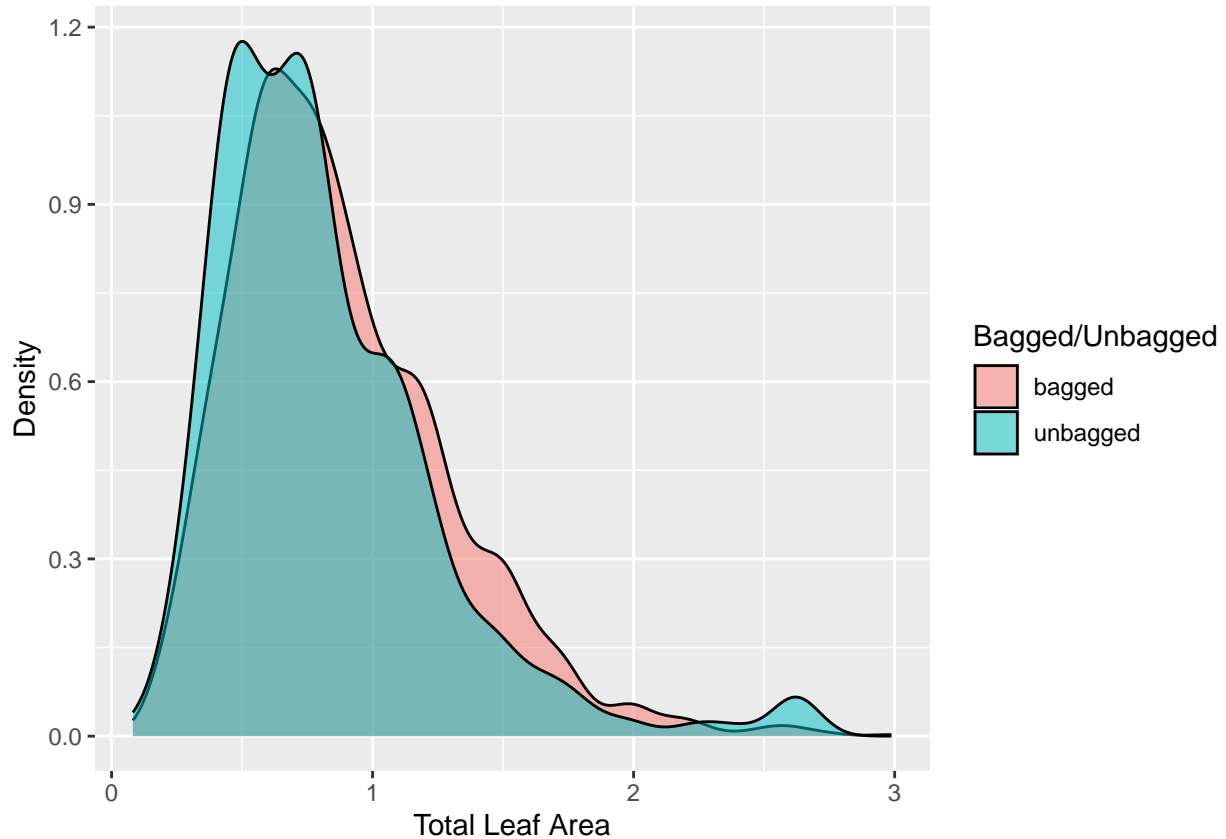
```
ggplot(Copy_of_Caterpillar, aes(birdtreatment)) + geom_bar() + xlab("Bagged/Unbagged") + ylab("Count")
```



- As shown in the bar graph, there are far more data points for bagged branches than for unbagged branches. There are around 2000 data points for the experimental bagged branches and less than 1000 for the control unbagged branches.

Bivariate Exploration

```
ggplot(Copy_of_Caterpillar, aes(x=totallvarea_m2, fill=birdtreatment)) + geom_density(alpha=.5) + xlab(
```



- The density graph shows how often the bagged or unbagged branches had the variety of total leaf areas. Bagged branches seem to have a larger density of total leaf areas in the 1-2 m² range, and unbagged branches has higher density of total leaf areas in the highest and lowest values of total leaf density.

Conclusion

- This data does not seem to support the idea that the caterpillars are able to consume more leaves on a bagged branch than a unbagged branch. It is difficult to say for sure whether the bagging of a branch aided in the caterpillar further reducing the total leaf area. Further mathematical analysis might be able to produce more evidence. Additionally, the experiment's methodology may not fully support the logic of the research question being asked.